Exp Appl Acarol (2011) 53:147–165
DOI 10.1007/s10493-010-9392-6

# Multiple transcripts encode glucose 6-phosphate dehydrogenase in the southern cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*

**Pia Untalan Olafson · Kevin B. Temeyer · John H. Pruett**

Received: 28 April 2010 / Accepted: 16 July 2010 / Published online: 15 August 2010


**Abstract** Glucose 6-phosphate dehydrogenase (G6PDH) is an enzyme that plays a critical role in the production of NADPH. Here we describe the identification of four transcripts (*G6PDH-A, -B, -C,* and *-D*) that putatively encode the enzyme in the southern cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. The genomic DNA that is spliced to produce *G6PDH-A* and *-B* is 8,600–9,000 bases in length and comprises 12 exons. Comparison of the *R. microplus G6PDH* gene structure with those available from insects and mammals revealed that the tick gene is most like that of humans. Detection of the four transcripts was evaluated by quantitative RT-PCR using template from larvae, unfed adult females and males, salivary gland tissues from 2- to 3-day-fed adult females and males, and salivary gland tissue of 4- to 5-day-fed adult females. The *G6PDH-A* and *-C* transcripts were present in all templates, and both displayed induced expression in salivary gland tissue of fed, adult females but not matched males. The *G6PDH-D* transcript was detected only in unfed adults and in larvae, a stage in which it was most abundant relative to the other three transcripts. The *G6PDH-B* transcript, while detectable in all templates, was of low copy number suggesting it is a rare transcript. Induced expression of *G6PDH-A* and *G6PDH-C* in fed females may play a role in the tolerance of oxidative stress that is induced upon feeding, and the transcript abundance in fed females may be a function of bloodmeal volume and the time adult females spend on the host relative to adult males.

**Keywords** G6PDH · Arthropod · Ixodid · Cattle tick

**Electronic supplementary material** The online version of this article (doi:10.1007/s10493-010-9392-6) contains supplementary material, which is available to authorized users.

P. U. Olafson (✉) · K. B. Temeyer · J. H. Pruett
USDA, Agricultural Research Service, Knipling-Bushland U. S. Livestock Insects Research Laboratory, 2700 Fredericksburg Rd., Kerrville, TX 78028, USA
e-mail: pia.olafson@ars.usda.gov

## Introduction

Glucose 6-phosphate dehydrogenase (G6PDH; EC 1.1.1.49) is an NADPH-producing enzyme that plays a critical role in the oxidative stage of the pentose phosphate pathway. NADPH is important in fatty acid synthesis and in detoxification reactions, acting as a cofactor for the regeneration of reduced glutathione and enabling activity of NADPH-dependent enzymes necessary for the protection of cells from oxidative damage (Stryer 1995). The G6PDH enzyme has been reported as playing a role in resistance to carcinogens (Yeh et al. 2001) and pyrethroids (Kumar et al. 1991), and deficiencies of the enzyme manifest as neonatal jaundice and hemolytic anemia in humans (Luzzatto 2006).

Blood-feeding arthropods ingest large volumes of blood and must adapt to the oxidative stress that is induced upon bloodmeal uptake (Graça-Souza et al. 2006). In the southern cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*, the breakdown of hemoglobin during the digestion of blood occurs in the digest cells of the midgut resulting in the release of heme (Lara et al. 2003). Free heme, toxic because of its potential to form oxygen radicals that promote oxidative damage to cells (Graça-Souza et al. 2006), remains intracellular and is subsequently detoxified by the tick in hemosomes, specialized detoxification organelles (Lara et al. 2005). While G6PDH plays an important biological role in protection of mammalian cells from oxidative damage, little is know about G6PDH in ticks (Gadallah et al. 1990; Moraes et al. 2007). Ticks are of veterinary and medical importance world-wide, and have a significant economic impact on the livestock industry with an estimated annual world-wide loss in the range of $18 billion USD (deCastro 1997). In addition, the spread of tick-borne diseases has had a dramatic effect on livestock morbidity and mortality, subsequently affecting the livelihood of farming communities in developing countries (Jongejan and Uilenberg 2004). Tick control methods have relied primarily on the use of acaricides, but this intensive use has resulted in the development of resistant tick populations (George et al. 2004). A more comprehensive understanding of tick biology would be beneficial to the design of alternative tick control technologies, including the identification of new control targets. Here, we describe four transcripts that encode G6PDH in this tick species and report on the partial genomic organization of *G6PDH* in *R. microplus*.

## Materials and methods

### Tick strains and tissues

Colony ticks of the Muñoz and San Román strains (Li et al. 2003) of *R. microplus* were obtained from the Cattle Fever Tick Research Laboratory (CFTRL) at Moore Air Base in Edinburg, Texas. The ticks were maintained as described by Davey et al. (1980). Unfed larvae between 12 and 16 days of age were collected and stored at −70°C prior to use. Unfed adult males and females were obtained by collecting meta-nymphs from a bovine host and allowing the specimens to molt into adults in chamber, after which they were stored at −70°C. Some adults were allowed to feed on a naïve bovine host for approximately 2–3 days (males and females) and 4–5 days (females only) post-nymphal molt prior to being removed from the host for dissection. Tick salivary glands were dissected at these timepoints and the tissues were stored in RNA*later*® (Ambion, Austin, TX) at −70°C prior to processing.

## Genomic DNA isolation, sequencing, and exon–intron characterization

A search of the *B. microplus* gene index, version 2.0 housed at the Dana Farber Cancer Institute site (http://compbio.dfci.harvard.edu/tgi/) for G6PDH, identified a tentative consensus sequence, TC67, the largest open reading frame of which displays the highest similarity at the amino acid level to G6PDH from *Rattus norvegicus*. Oligonucleotide primers (Table 1) designed approximately every 300–500 bases along the TC67 transcript were utilized to amplify overlapping regions of pooled, genomic DNA (gDNA) in order to identify the location of possible introns within the putative *B. microplus* G6PDH coding sequence. Genomic DNA was isolated from 0.3 g of tick larvae (Muñoz strain, $F_{24;}$ San Román strain, $F_{31}$) utilizing the method described by Hill and Gutierrez (2003). Amplified products were cloned into the pCR®4-TOPO® plasmid vector via the TOPO-TA Cloning® Kit (Invitrogen, Calrsbad, CA). Bacterial colony isolates from the cloning reactions were used to inoculate liquid bacterial media supplemented with appropriate antibiotic, and plasmid DNA was isolated from these cultures using the QIAprep® Spin Miniprep Kit (Qiagen, Valencia, CA). Plasmid DNA from five isolates was sequenced using BigDye™ chemistry (Applied Biosystems, Foster City, CA) with universal M13F and M13R primers, and subsequently analyzed on an ABI-3130xl instrument (Applied Biosystems). Since the cloned inserts were >1.5 kb in length, it was necessary to design additional primers to enable full-length sequencing of a single representative plasmid clone. Intron sequences were located by comparing the gDNA and mRNA/cDNA sequence using the Spidey algorithm available at the National Center for Biotechnology Information (NCBI) website [http://www.ncbi.nlm.nih.gov].

Approximately 4,100 bases of genomic DNA sequence upstream of the TC67 start codon was obtained using the GenomeWalker™ Universal Kit (Clontech, Mountain View, CA) following the manufacturer's protocol and primers designed from accumulated sequence data (Table 1, primers labeled 'GW'). This system was also used to obtain an additional 580 bases of sequence downstream of the putative translational stop codon using a primer designed from within the last exon of the gene. Fragments of interest were cloned and sequenced, as described above. In parallel, an expressed sequence tag (EST) putatively encoding G6PDH, denoted G6PDH-C, was identified from an *R. microplus* salivary gland EST library (Olafson, unpublished). The entire length of the clone was sequenced utilizing primers designed from available clone sequence data.

## gDNA sequence from an individual larva

Since the sequence data obtained was from template of a pooled origin, hence a 'mosaic' gene sequence, we used a combination of primers spanning the accumulated sequence to amplify gDNA from an individual tick larva to validate the sequence information. A single larva (Muñoz strain, $F_{24}$) was placed in a 0.5 ml homogenizer tube and macerated using a liquid-$N_2$ cooled disposable pestle. One-hundred µl of homogenizing buffer (10 mM Tris–HCl, pH8.3, 50 mM potassium chloride) was added, and the tick further homogenized prior to boiling for 10 min. The sample was microcentrifuged at 20,000×*g*, 4°C for 5 min, and the supernatant was utilized as template in PCR amplification, the products of which were cloned and sequenced as above. Overlapping segments of the gene were aligned to produce the presented gene sequence.

**Table 1** Oligonucleotide primers utilized to characterize *R. microplus G6PDH*

| Primer name | Sequence (5′ → 3′)[a] | DNA strand | Location[b] | $T_{annealing}$ (RT-PCR) (°C) |
|---|---|---|---|---|
| G6PDH-F1 | ***ATG***CCTCAAACATCGTTCC | Sense | 1–19 of G6PDH-A transcript; 5,311–5,329 | 65 |
| G6PDH-F1b | ***ATG***AGAAGAAACGAAGGGGCACTG | Sense | 1–24 of G6PDH-B transcript; 6,289–6,312 | 66 |
| G6PDH-F1c | ***ATG***GCTGGCCGCGCGGCGAGTGAA | Sense | 1–24 of G6PDH-C transcript | 64 |
| G6PDH-F1d | ***ATG***GTCAAGTCCACGGGA | Sense | 1–18 of G6PDH-D transcript; 52–69 | 65 |
| G6PDH-R5 | ***TTA***AGCTTTAGTCCAACGGTATGTG | Anti-sense | Translation stop of G6PDH-A/-B/-C/-D; 13,733–13,709 | |
| G6PDH-F9 | ACCCTACGAACGACGGACTA | Sense | 5,664–5,683 | |
| G6PDH-F12 | TAAGCGCTTGAAGCCATTCT | Sense | 6,038–6,057 | |
| G6PDH-F19 | GCTCGGAACTCATTACTGTGG | Sense | 6,865–6,885 | |
| G6PDH-F2 | CAGCAACATCAAGCAGCATT | Sense | 7,259–7,278 | |
| G6PDH-F7 | AGTGTATTTGGGAGATGTTCATAGT | Sense | 7,526–7,550 | |
| G6PDH-F3 | TTATGCAGAACCACCTGCTGC | Sense | 7,913–7,933 | |
| G6PDH-F11 | GCCGCGCTATTTGAAGAGT | Sense | 7,927–79,245 | |
| G6PDH-F13 | AAGTTAACGGACAGGGAAGACA | Sense | 8,313–8,334 | |
| G6PDH-F20 | TCCTGGGATTGAGTTGAACC | Sense | 9,101–9,120 | |
| G6PDH-F25 | CCGGGATGTAGTCAGTTGCT | Sense | 10,211–10,230 | |
| G6PDH-F27 | TCAGATTTGGGTGCACGTTA | Sense | 10,691–10,710 | |
| G6PDH-F26 | GACCATCTGCACTGCTTGAA | Sense | 11,427–11,446 | |
| G6PDH-F4 | CAAGGCACTGAATGAGCGTA | Sense | 11,911–11,930 | |
| G6PDH-F5 | GCCTTCGACATTGAGGAGAC | Sense | 12,066–12,085 | |
| G6PDH-F10 | GGAACCAAATGTACTGAAAGGTT | Sense | 12,233–12,255 | |
| G6PDH-F14 | GCCGTGCTTGGTCATGTA | Sense | 12,648–12,665 | |
| G6PDH-R6 | CATGCAGGCACACCGAAG | Anti-sense | 5,765–5,748 | |
| G6PDH-R1 | AATGCTGCTTGATGTTGCTG | Anti-sense | 7,278–7,259 | |
| G6PDH-R2 | TCTGCATTTGTGGACACTGG | Anti-sense | 9,870–9,851 | |
| G6PDH-R8 | TTGTTGAAAATTGCAAGGAG | Anti-sense | 12,289–12,270 | |

**Table 1** continued

| Primer name | Sequence (5′ → 3′)[a] | DNA strand | Location[b] | $T_{annealing}$ (RT-PCR) (°C) |
|---|---|---|---|---|
| G6PDH-R4 | ACTGGTTTGGGCTTCTCCTT | Anti-sense | 13,182–13,201 | |
| D-F2 | AAACGCGCTCCTCTGGAATT | Sense | 34–53 of G6PDH-D 5′ UTR, Fig. 1 | |
| D-F1 | CACAGAAAGCACAGGTGCGC | Sense | 81–100 of G6PDH-D 5′ UTR, Fig. 1 | |
| Intron 1-R1 | AGCGCTTACCATAACGTTGG | Anti-sense | 6,026–6,045 | |
| Exon 1-Rev GW | AGATGTGTTGCTGTCCCTCCTGCACA | Anti-sense | 5,394–5,419 | |
| GW12 | TCCGCCTTTCATCAACCAACTACGTCAA | Anti-sense | 4,029–4,056 | |
| GW13 | ACATATCCGAAAACCCCCGTCAAGAACA | Anti-sense | 2,838–2,811 | |
| GW14 | CCAATGACGTGGCCCCTTTCTATGACTA | Anti-sense | 2,654–2,681 | |
| GW-Down2 | AGAAGCCCAAACCAGTGCCTTACGAGTA | Sense | 13,186–13,213 | |

[a] The translational start and stop for appropriate primers is identified by bold, italicized text

[b] Numbering based on pooled genomic DNA sequence (**DQ118973**)

Rapid amplification of cDNA ends (RACE)

RACE-Ready cDNAs (5′ and 3′) were synthesized utilizing the SMART™ RACE cDNA Amplification Kit (Clontech), as described in the manufacturer's protocol. The mRNA template for these reactions was isolated from unfed tick larvae and salivary glands from fed adult females using the Micro-FastTrack™ 2.0 kit (Invitrogen). The 5′RACE reaction for each template was directed using the G6PDH-R2 primer in a primary round of cycling followed by a nested reaction using the G6PDH-R1 primer, both in combination with a commercial adapter primer. Products of 3′ RACE were obtained utilizing G6PDH-F1, G6PDH-F1b, G6PDH-F1c, and G6PDH-F1d (Table 1) in combination with the adapter primers supplied by the manufacturer. Amplified products of interest were cloned into the pCR®4-TOPO® vector and five plasmid clones from each reaction were sequenced as described above.

RT-PCR and real time quantitative RT-PCR (q-RTPCR) to evaluate expression of transcripts

Total RNA was isolated from twenty larvae, twenty each of unfed adult males and females, and 20 pair each of salivary glands from 2 to 3 days fed, adult males and 2–3 days and 4–5 days fed, adult females utilizing RNAwiz™ Reagent (Ambion). Total RNA was DNAse treated using the DNA-*free*™ system (Ambion) prior to first-strand cDNA synthesis. First strand cDNA was synthesized using equal amounts of total RNA from each source (1.5 μg), 2 pmol G6PDH-R5 primer, and 0.5 mM dNTP mix that were denatured at 65°C for 5 min and then combined with 1 × First Strand Buffer (Invitrogen), 5 mM DTT, 40 U RNaseOUT™ (Invitrogen), and 200 U of SuperScript™ III Reverse Transcriptase (Invitrogen) in a total volume of 20 μl. The cDNAs were synthesized at 55°C for 1 h. The G6PDH-R5 primer used to direct cDNA synthesis contains the translational stop codon that is shared by the four transcripts being investigated.

Separate, hemi-nested RT-PCR amplifications were conducted utilizing primers G6PDH-F1, G6PDH-F1b, G6PDH-F1c, or G6PDH-F1d in combination with G6PDH-R5 (Table 1). First-strand cDNAs (larval, unfed adults, and salivary glands from fed adults) were used as template in 20 μl reactions consisting of 40 mM Tricine-KOH, 15 mM potassium acetate, 3.5 mM magnesium acetate, 3.75 μg/ml BSA, 0.005% Tween 20, 0.005% Nonidet-P40, 0.2 mM dNTPs (Applied Biosystems), 5 pmol each primer, and 1× Advantage® 2 Polymerase Mix (Clontech). Amplifications were conducted using an Eppendorf® Mastercycler® Gradient thermal cycler (Brinkmann, Westbury, NY), and cycling conditions for the four primer pairs (listed above) consisted of an initial denaturation at 95°C for 3 min followed by 30 cycles of 95°C for 30 s, optimal primer pair annealing temperature (Table 1) for 30 s and 68°C for 90 s with a final extension of 68°C for 3 min. Products were resolved by agarose gel electrophoresis and visualized by staining with GelStar® Nucleic Acid Gel stain (Cambrex, Rockland, ME). The amplified G6PDH-A, -B, -C, and -D products of each relevant template, when present, were cloned into the pCR®4-TOPO® vector and sequenced with M13-F and -R primers to verify insert sequences. Two representative clones of each transcript from the respective templates were sequenced full-length using the internal G6PDH-F2, G6PDH-F3, and G6PDH-F4 primers in order to evaluate sequence polymorphisms.

To determine absolute quantities of G6PDH mRNA expression in larval, unfed adult, and two timepoints of fed adults, oligonucleotide primer pairs were designed such that the forward primer was located in a region unique to the G6PDH-A, -B-, -C, or -D transcript

**Table 2** Oligonucleotide primers used in quantitative RT-PCR

| Primer pair | Target transcript | Position[a] | Sequence (5′ → 3′) | Product size | $T_{annealing}$ (°C) |
|---|---|---|---|---|---|
| A13-rt.Fwd | A | 13 → 30 | TCGTTCCCGCCTTCTCTC | 251 | 58 |
| A263-rt.Rev | | 246 ← 263 | ACAATCTTCGCCCACAGC | | |
| B2-rt.Fwd | B | 2 → 25 | TGAGAAGAAACGAAGGGGCACTG | 75 | 54 |
| B76-rt.Rev | | 56 ← 76 | TGCGAGCATAACCAACGAAC | | |
| C74-rt.Fwd | C | 74 → 93 | GCTCCTACTCCAGTGAAGAC | 100 | 60 |
| C173-rt.Rev | | 154 ← 173 | ACGACGAAGATGTGTTGTTG | | |
| D3-rt.Fwd | D | 3 → 20 | CAGAGTATGGTCAAGTCC | 99 | 53 |
| D101-rt.Rev | | 82 ← 101 | GAAGATGTGTTGTTGTCC | | |

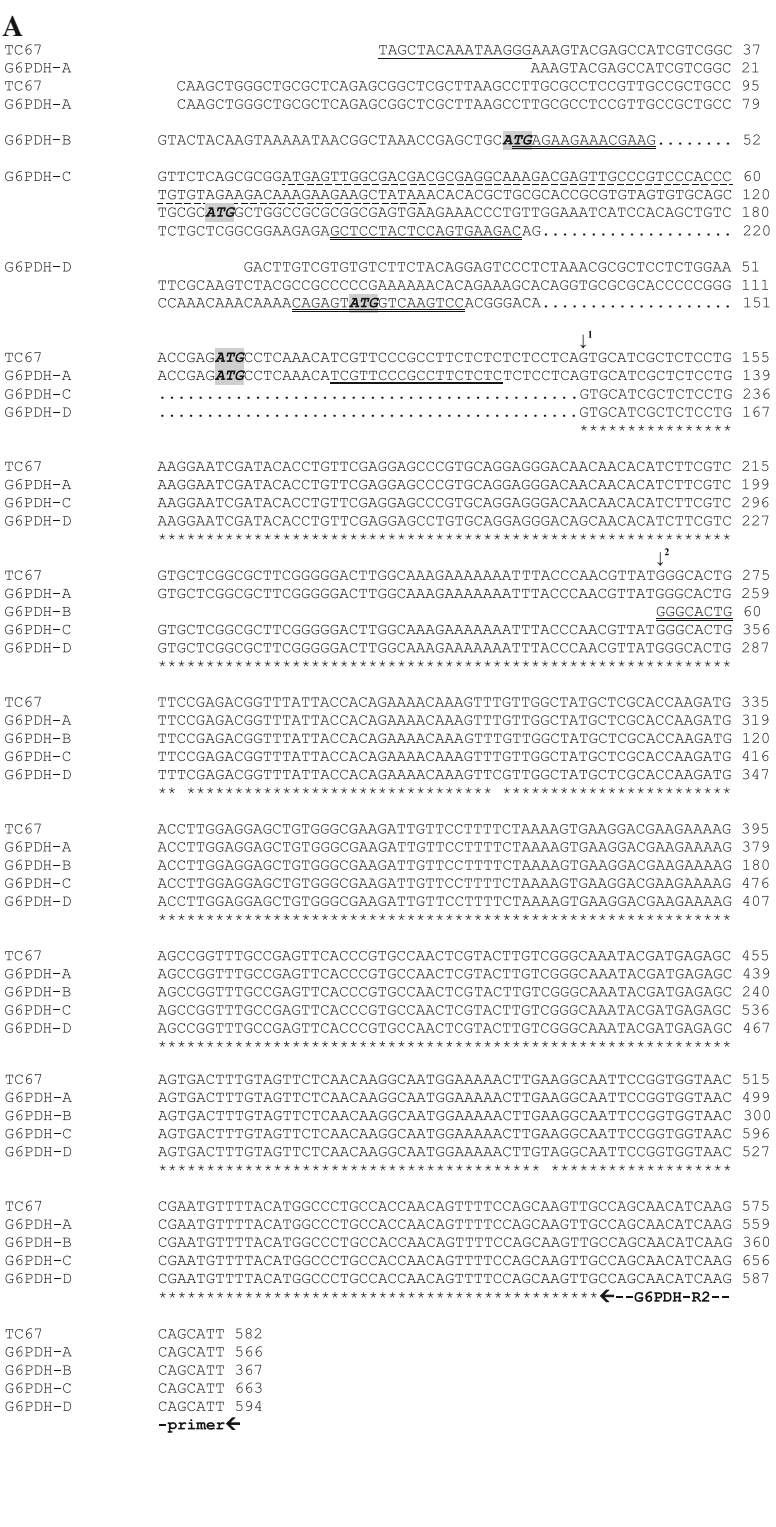
[a] Position corresponds to location within the relevant transcript sequences. Forward primer sequences are indicated in Fig. 1

and the reverse primer was located in a region common to all four (Table 2, Fig. 1). In all cases, the primers would amplify a substantially different sized product should gDNA contamination be present. Twenty-five µl amplification reactions contained 1× iTaq SYBR Green SuperMix with ROX (Bio-Rad Laboratories, Hercules, CA), 250 nM each of forward and reverse primer, and 50 ng of G6PDH-R5 primed cDNA (synthesis described above). Evaluation of the G6PDH-B transcript utilized 100 ng of cDNA template. Standard curves for each primer set were constructed with reaction mixtures containing serially diluted, linearized plasmids ($10^0$ to $10^8$ target copies) with transcript-specific inserts. The standard curve correlation coefficients for each run ranged from 0.990 to 0.999, with slopes that ranged from −3.57 to −3.67. Standard curve and experimental templates for the respective transcripts were analyzed in triplicate on the same plate, thus exposed to identical cycling conditions. Cycling, conducted on an ABI Prism® 7000 Sequence Detection System (Applied Biosystems), was as follows: initial denaturation at 95°C for 2 min 30 s, and 35 cycles of 95°C for 30 s, optimal annealing temperature (see Table 2) for 60 s, and 72°C for 30 s. A dissociation protocol was included to verify the presence of a single melting peak (single product) for each experiment. The copy number of each transcript/50 ng cDNA template was determined by calculating the mean of the threshold cycle value for the experimental templates plotted against the appropriate standard curve using the Sequence Detection System (SDS), version 1.1 software package (Applied Biosystems).

## DNA and amino acid sequence analysis

The ClustalW executable program (Thompson et al. 1994) was utilized for sequence assembly. The genomic sequence for G6PDH was overlayed on the respective G6PDH cDNA sequences from several organisms to evaluate intron location and aid in

**Fig. 1** Nucleotide sequence of 5′RACE products for four *R. microplus* transcripts, G6PDH-A, -B, -C, and -D. **a** The translation initiation codon for each is designated by ***ATG***, and the G6PDH-R2 primer used to obtain each fragment is identified. $\downarrow^{1}$ denotes the base at which the G6PDH-A, -C, and -D transcripts begin to align, and $\downarrow^{2}$ denotes the base at which G6PDH-B starts to align with the other three transcripts. The location of the forward primer used for q-RTPCR that is unique to the respective transcripts is double-underlined (=). A 75 bp untranslated open reading in the 5′ UTR of *G6PDH-C* is identified by a thatched underline (- - - -). **b** The variant 3′ UTR-short and -long sequences identified by 3′RACE are presented. The translation stop codon (***TAA***), and the non-canonical and canonical polyadenylation signals, **TATAAA** and [***AATAAA***], respectively, are identified


A
TC67 TAGCTACAAATAAGGGAAAGTACGAGCCATCGTCGGC 37
G6PDH-A AAAGTACGAGCCATCGTCGGC 21
TC67 CAAGCTGGGCTGCGCTCAGAGCGGCTCGCTTAAGCCTTGCGCCTCCGTTGCCGCTGCC 95
G6PDH-A CAAGCTGGGCTGCGCTCAGAGCGGCTCGCTTAAGCCTTGCGCCTCCGTTGCCGCTGCC 79
G6PDH-B GTACTACAAGTAAAAATAACGGCTAAACCGAGCTGCATGAGAAGAAACGAAG..... 52
G6PDH-C GTTCTCAGCGCGGATGAGTTGGCGACGACGCGAGGCAAAGACGAGTTGCCCGTCCCACCC 60
TGTGTAGAAGACAAAGAAGAAGCTATAAACACACGCTGCGCACCGCGTGTAGTGTGCAGC 120
TGCGCCATGGCTGGCCGCGCGGCGAGTGAAGAAACCCTGTTGGAAATCATCCACAGCTGTC 180
TCTGCTCGGCGGAAGAGAGCTCCTACTCCAGTGAAGACAG..... 220
G6PDH-D GACTTGTCGTGTGTCTTCTACAGGAGTCCCTCTAAACGCGCTCCTCTGGAA 51
TTCGCAAGTCTACGCCGCCCCCGAAAAAACACAGAAAGCACAGGTGCGCGCACCCCCGGG 111
CCAAACAAACAAAACAGAGTATGGTCAAGTCCACGGGACA..... 151
TC67 ACCGAGATGCCTCAAACATCGTTCCCGCCTTCTCTCTCTCCTCAGTGCATCGCTCTCCTG 155
G6PDH-A ACCGAGATGCCTCAAACATCGTTCCCGCCTTCTCTCTCAGTGCATCGCTCTCCTG 139
G6PDH-C .....GTGCATCGCTCTCCTG 236
G6PDH-D .....GTGCATCGCTCTCCTG 167
TC67 AAGGAATCGATACACCTGTTCGAGGAGCCCGTGCAGGAGGGACAACAACACATCTTCGTC 215
G6PDH-A AAGGAATCGATACACCTGTTCGAGGAGCCCGTGCAGGAGGGACAACAACACATCTTCGTC 199
G6PDH-C AAGGAATCGATACACCTGTTCGAGGAGCCCGTGCAGGAGGGACAACAACACATCTTCGTC 296
G6PDH-D AAGGAATCGATACACCTGTTCGAGGAGCCTGTGCAGGAGGGACAGCAACACATCTTCGTC 227
TC67 GTGCTCGGCGCTTCGGGGGACTTGGCAAAGAAAAAAATTTACCCAACGTTATGGGCACTG 275
G6PDH-A GTGCTCGGCGCTTCGGGGGACTTGGCAAAGAAAAAAATTTACCCAACGTTATGGGCACTG 259
G6PDH-B GGGCACTG 60
G6PDH-C GTGCTCGGCGCTTCGGGGGACTTGGCAAAGAAAAAAATTTACCCAACGTTATGGGCACTG 356
G6PDH-D GTGCTCGGCGCTTCGGGGGACTTGGCAAAGAAAAAAATTTACCCAACGTTATGGGCACTG 287
TC67 TTCCGAGACGGTTTATTACCACAGAAAACAAAGTTTGTTGGCTATGCTCGCACCAAGATG 335
G6PDH-A TTCCGAGACGGTTTATTACCACAGAAAACAAAGTTTGTTGGCTATGCTCGCACCAAGATG 319
G6PDH-B TTCCGAGACGGTTTATTACCACAGAAAACAAAGTTTGTTGGCTATGCTCGCACCAAGATG 120
G6PDH-C TTCCGAGACGGTTTATTACCACAGAAAACAAAGTTTGTTGGCTATGCTCGCACCAAGATG 416
G6PDH-D TTTCGAGACGGTTTATTACCACAGAAAACAAAGTTCGTTGGCTATGCTCGCACCAAGATG 347
TC67 ACCTTGGAGGAGCTGTGGGCGAAGATTGTTCCTTTTCTAAAAGTGAAGGACGAAGAAAAG 395
G6PDH-A ACCTTGGAGGAGCTGTGGGCGAAGATTGTTCCTTTTCTAAAAGTGAAGGACGAAGAAAAG 379
G6PDH-B ACCTTGGAGGAGCTGTGGGCGAAGATTGTTCCTTTTCTAAAAGTGAAGGACGAAGAAAAG 180
G6PDH-C ACCTTGGAGGAGCTGTGGGCGAAGATTGTTCCTTTTCTAAAAGTGAAGGACGAAGAAAAG 476
G6PDH-D ACCTTGGAGGAGCTGTGGGCGAAGATTGTTCCTTTTCTAAAAGTGAAGGACGAAGAAAAG 407
TC67 AGCCGGTTTGCCGAGTTCACCCGTGCCAACTCGTACTTGTCGGGCAAATACGATGAGAGC 455
G6PDH-A AGCCGGTTTGCCGAGTTCACCCGTGCCAACTCGTACTTGTCGGGCAAATACGATGAGAGC 439
G6PDH-B AGCCGGTTTGCCGAGTTCACCCGTGCCAACTCGTACTTGTCGGGCAAATACGATGAGAGC 240
G6PDH-C AGCCGGTTTGCCGAGTTCACCCGTGCCAACTCGTACTTGTCGGGCAAATACGATGAGAGC 536
G6PDH-D AGCCGGTTTGCCGAGTTCACCCGTGCCAACTCGTACTTGTCGGGCAAATACGATGAGAGC 467
TC67 AGTGACTTTGTAGTTCTCAACAAGGCAATGGAAAAACTTGAAGGCAATTCCGGTGGTAAC 515
G6PDH-A AGTGACTTTGTAGTTCTCAACAAGGCAATGGAAAAACTTGAAGGCAATTCCGGTGGTAAC 499
G6PDH-B AGTGACTTTGTAGTTCTCAACAAGGCAATGGAAAAACTTGAAGGCAATTCCGGTGGTAAC 300
G6PDH-C AGTGACTTTGTAGTTCTCAACAAGGCAATGGAAAAACTTGAAGGCAATTCCGGTGGTAAC 596
G6PDH-D AGTGACTTTGTAGTTCTCAACAAGGCAATGGAAAAACTTGTAGGCAATTCCGGTGGTAAC 527
TC67 CGAATGTTTTACATGGCCCTGCCACCAACAGTTTTCCAGCAAGTTGCCAGCAACATCAAG 575
G6PDH-A CGAATGTTTTACATGGCCCTGCCACCAACAGTTTTCCAGCAAGTTGCCAGCAACATCAAG 559
G6PDH-B CGAATGTTTTACATGGCCCTGCCACCAACAGTTTTCCAGCAAGTTGCCAGCAACATCAAG 360
G6PDH-C CGAATGTTTTACATGGCCCTGCCACCAACAGTTTTCCAGCAAGTTGCCAGCAACATCAAG 656
G6PDH-D CGAATGTTTTACATGGCCCTGCCACCAACAGTTTTCCAGCAAGTTGCCAGCAACATCAAG 587
G6PDH-R2--
TC67 CAGCATT 582
G6PDH-A CAGCATT 566
G6PDH-B CAGCATT 367
G6PDH-C CAGCATT 663
G6PDH-D CAGCATT 594
-primer

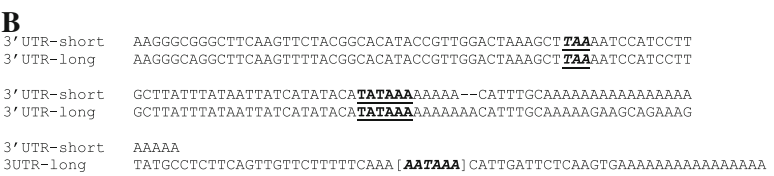



**Fig. 1** continued

construction of Fig. 4. The relevant Genbank accession numbers are: *R. norvegicus*-protein: **NP_05872**; *Drosophila melanogaster*—gDNA: **AE014298**, cDNA: **NM_078687** and **NM_167676**; *Ips typographus*—gDNA/cDNA: **AY845218**; *Homo sapiens*—gDNA: **NT_011726**, cDNA: **NM_000402** and **NM_001042351**. Gene structures for *Macropus robustus* (Loebel and Johnston 1997) and *Takifugu rubripes* (Mason et al. 1995) *G6PDH* were obtained from the literature. Conserved amino acid resides were identified using the Conserved Domain Database at the NCBI site.

## Results

### Four transcripts encode G6PDH in *Rhipicephalus microplus*

A full-length cDNA sequence putatively encoding *R. microplus* G6PDH [BmiGI (Guerrero et al. 2005), sequence TC67, 1,745 bases], was utilized as a starting point to characterize the *G6PDH* gene sequence from this organism. The TC67 transcript encoded an open reading frame of 1,548 bp that displayed similarity to G6PDH sequences from various *Drosophila* sp. ($E = 2e^{-15}$) and its conceptual translation (515 aa) encoded the conserved $NADP^+$ binding domain and C-terminal domain that are characteristic of G6PDH.

Various oligonucleotide primers (Table 1) were designed to amplify overlapping segments that span the length of the open reading frame in order to identify the location of intron sequences and to characterize the 5′ and 3′ UTRs of the transcript. While characterizing the 5′ UTR of the transcript using cDNA template prepared from pooled larvae (L) or pooled salivary glands of fed adult females (SG), several amplified products approx. 550–600 bases in length were observed along with faintly staining fragments that were 300–500 bases in length (data not shown). Sequence analysis of two fragments from the L reaction (594 bp and 566 bp) and one fragment from the SG reaction (367 bp) revealed they differed substantially from each other at their 5′ ends (Fig. 1). Comparison with the corresponding region of the TC67 sequence revealed that the transcript with the 566 bp 5′end, designated G6PDH-A (Genbank Accession No. **EU595878**), displayed 100% similarity to TC67; albeit, the -A transcript was 16 bases shorter than that of TC67. The transcript with the 367 bp 5′ end, designated G6PDH-B (Genbank Accession No. **EU595879**), was 86% similar to TC67 while the transcript dictated by the largest 5′RACE fragment (594 bp), designated G6PDH-D (Genbank Accession No. **EU595880**), displayed 75% sequence similarity to TC67. In parallel, a transcript encoding a putative G6PDH, designated G6PDH-C (Genbank Accession No. **EF411191**), was identified from an *R. microplus* salivary gland EST library (Olafson, unpublished), and was 76% similar to TC67. The 5′RACE sequences presented in Fig. 1A depict these varied ends, and highlight the dissimilarity between G6PDH-A/TC67 and the other three transcripts. The nucleotide position at which the G6PDH-A, -C, and -D transcripts begin aligning with each other are

identified in Fig. 1a (indicated by $\downarrow^1$) along with the nucleotide position at which the G6PDH-B transcript starts to align with these three (indicated by $\downarrow^2$).

Analysis of the 3′RACE clones obtained for the four putative G6PDH transcripts consistently produced two fragments that differed in size by ~75 bases, one designated 3′ UTR-short and the other designated 3′ UTR-long, the latter of which contained a canonical polyadenylation signal (AAUAAA; Fig. 1b).

## Transcription profile of G6PDH-A, -B, -C, and -D

Representation of the four G6PDH transcripts in pooled larvae, unfed adult males or females, and salivary glands of fed adult males (2–3 days) or fed adult females (2–3 days and 4–5 days) was evaluated using real-time quantitative reverse transcriptase PCR (qRT-PCR, Table 3). In an effort to obtain a more accurate reflection of transcript presence, first strand cDNA synthesized from these stages and tissues was directed using the G6PDH-R5 primer that is common to all transcripts and was designed to include the translational stop codon.

The G6PDH-A transcript was detected in all templates evaluated, with equivalent amounts in both sexes of unfed adults. The transcript was reduced eightfold in 2 days fed males relative to unfed males and ca. threefold in 2 days fed females relative to unfed females, but was upregulated sixfold in 4 days fed females relative to unfed females. As with G6PDH-A, the -C transcript was detected in all templates. Equivalent amounts of transcript were observed in unfed and fed adult males, and a significant upregulation in number of transcripts from 1.4-fold to 12-fold was noted in 2 days fed to 4 days fed females, respectively, relative to unfed females. Interestingly, the G6PDH-D transcript was detected most abundantly in larvae compared with the other three transcripts and was present in equivalent amounts in unfed males and females, but was not detected in fed males or fed females. Although the G6PDH-B transcript was present in all templates analyzed, the number of copies detected suggests it is a rare transcript.

## Single nucleotide polymorphisms within the G6PDH-A, -B, -C, and -D transcripts

The G6PDH-R5 primed cDNAs were used as template in a semi-nested reaction with oligonucleotide primers corresponding to the translational start and stop regions of G6PDH-A, -B, -C, and -D (Table 1). All of the amplified products that were detectable by

**Table 3** Transcription profile of *G6PDH-A*, *-B*, *-C*, *and -D* analyzed by q-RTPCR[a]

| Template | G6PDH-A | G6PDH-B | G6PDH-C | G6PDH-D |
|---|---|---|---|---|
| Larvae | 1,212.48 ± 58.86 | 10.28 ± 2.65 | 256.1 ± 2.71 | 1,851.05 ± 60.04 |
| Unfed Adult ♂ | 532.23 ± 12.73 | 9.72 ± 1.08 | 58.59 ± 3.36 | 50.37 ± 5.70 |
| Fed Adult ♂ (2–3 days) | 66.24 ± 7.64 | 4.44 ± 0.13 | 57.07 ± 2.56 | – |
| Unfed Adult ♀ | 542.28 ± 13.10 | 5.02 ± 0.74 | 160.53 ± 16.58 | 53.04 ± 7.20 |
| Fed Adult ♀ (2–3 days) | 202.10 ± 10.56 | 4.17 ± 0.28 | 229.59 ± 3.77 | – |
| Fed Adult ♀ (4–5 days) | 3,327.09 ± 116.64 | 20.25 ± 2.26 | 1,940.19 ± 124.36 | – |

[a] Quantity is presented as mRNA copies/unit where 1 unit = 50 ng cDNA. For G6PDH-B, 100 ng cDNA was used as template and the quantities reported were adjusted accordingly to be consistent with the other transcripts. The values presented are mean ± SE

RT-PCR were subsequently cloned and sequenced to evaluate single nucleotide polymorphic sites within the transcripts. Full-length sequence information for two isolates from each cloning reaction were aligned for each of the relevant G6PDH-A, -B, -C, and -D transcripts (Supp. Figs. 1 thru 4). The positions of single nucleotide differences present in 2 or more of the aligned G6PDH-A transcripts were noted and then pairwise compared with the aligned -B, -C, and -D transcripts to identify shared and unique differences, summarized in Table 4. Twenty polymorphic (shared) sites and twelve unique sites were identified in the 12 G6PDH-A transcripts, while 16 polymorphic positions each were identified within the 10 G6PDH-B, the 11 G6PDH-C, and the six G6PDH-D transcripts evaluated. Transition substitutions were observed at all but one site, position 1,122 of G6PDH-A. A comparison among the G6PDH-A, -B, -C, and -D transcripts revealed four polymorphic sites that were consistently present across the 39 total transcripts analyzed and were located at positions 381, 819, 843, and 933 relative to the G6PDH-A transcript. Three sites were unique to the G6PDH-B and -D transcripts: positions 910, 1,005, and 1,344, the latter of which was also polymorphic within the ESTs aligned to create the TC67 transcript (Dana Farber Cancer Institute site, data not shown). The polymorphism at position 1,029 (Table 4) was of interest primarily because three nucleotides were observed among the G6PDH-A and -D transcripts at this location, suggesting that there may be greater than one copy of the *G6PDH* gene within the haploid genome.

Alignment of the conceptually translated products from *R. microplus* with G6PDH from *D. melanogaster,* human, rat, and the bark beetle revealed the conservation of functional domains that are characteristic of G6PDH molecules (Fig. 2). The alignment in Fig. 2 is presented to demonstrate the amino acid differences at the 5′ ends of the conceptual translations, and does not take into account the various SNPs observed within the transcripts.

## Partial genomic characterization of *Rhipicephalus microplus G6PDH*

The gDNA sequence that is spliced to produce the *R. microplus G6PDH-A* and *-B* transcripts, including sequence information downstream of the translational stop, was initially obtained from pooled gDNA template and comprised ~9,000 bases. Due to the heterogeneous nature of pooled gDNA, the same region was analyzed from an individual larva (M1) and was ~8,600 bases in length, a difference that is attributable to a number of short insertion/deletion (indel) events and two indels, 167 bases and 254 bases ('indel 254') in length, that were present in the pooled sequence (Supp. Figs. 5 and 6).

The 5′ region of the G6PDH-D transcript was not present within these 8,600–9,000 bases of genomic sequence; thus, a primer walking approach was initiated upstream of the available data and ~4,100 bases were accumulated following three rounds of primer walking (Fig. 3). Separately, a primer designed from within the 5′-UTR region of G6PDH-D (D-F2) was used in conjunction with a primer designed from within a defined intron (Intron 1-R1) to amplify pooled gDNA (Fig. 3). This product was then used as template in a nested amplification with primers D-F1 and Zw-GW13, producing a 2,900 base pair fragment that included the G6PDH-D start sequence and overlapped the primer walking segment by 1,420 bases (Fig. 3). The accumulated genomic sequence obtained from pooled template was 14,311 bases in length (Genbank Accession Number **DQ118973**, Supp. Fig. 5).

Primers designed within this upstream 5,300 base region were used to evaluate gDNA from individual M1 and identified an 'indel 254' sequence as well as a 672 base indel that were not within the sequence from pooled gDNA. Conversely, an 800 base indel present

**Table 4** Summary of single nucleotide polymorphic positions identified within the G6PDH-A, -B, -C, and -D transcripts

| Position[a] | A | B | C | D |
|---|---|---|---|---|
| 84 | • | | | |
| 99 | • | | | • |
| 177 | | • | | • |
| 210 | • | • | • | |
| 303 | • | | • | • |
| 312 | | | | • |
| 360 | • | | • | • |
| 381 | • | • | • | • |
| 387 | • | | | |
| 414 | • | | | |
| 417 | • | | | |
| 549 | • | | | |
| 600 | • | • | • | |
| 624 | • | | • | • |
| 693 | • | • | • | |
| 786 | • | | | |
| 816 | • | • | | • |
| 819 | • | • | • | • |
| 843 | • | • | • | • |
| 855 | • | | | |
| 879 | • | • | | • |
| 882 | • | • | • | |
| 897 | • | | | |
| 910 | | • | | • |
| 918 | • | | • | • |
| 933 | • | • | • | |
| 944 | • | | | |
| 959 | • | • | • | |
| 1,005 | | • | | • |
| 1,029[b] $_{3\ nts}$ | • | | | • |
| 1,104 | • | | | |
| 1,122[b] $_{tv}$ | • | | | |
| 1,323 | • | | • | • |
| 1344[b] $_{TIGR}$ | | • | | • |
| 1,404 | • | • | • | |
| 1,506 | • | | | |
| 1,518 | • | • | • | |

[a] Nucleotide positions listed are based on the G6PDH-A open reading frame sequence (ATG → TAA)

[b] *3 nts* position at which three nucleotides were observed when multiple transcripts were sequenced (see text); *tv* transversion polymorphism; *TIGR* position at which a polymorphism was identified between the transcripts used by TIGR to create the TC67 sequence

within this region of the pooled sequence was not present in M1 (Supp. Figs. 5 and 6). While a majority of the pooled sequence was validated using individual M1, the first 840 bases of sequence from pooled template that included the G6PDH-D start could not be amplified from this particular individual (Supp. Figs. 5 and 6). To identify whether this was a polymorphism within M1, the same region was amplified using gDNA from another individual larva, M2, with the same result. Thus, the gene structure reported here is based

```
G6PDH-A                                MPQTSFPPSLSPQCIALLKESIHLFEEPVQEGQQHIFVVLG 41
G6PDH-B                                                        MRRNEG  6
G6PDH-C          MAGRAASEETLLEIIHSCLCSAEESSYSSEDRCIALLKESIHLFEEPVQEGQQHIFVVLG 60
G6PDH-D                                MVKSTGQCIALLKESIHLFEEPVQEGQQHIFAVLG 35
Rn-515                                MAEQVALSRTQVCGILREELYQGDAFHQADTHIFIIMG 38
Hs.B-515                              MAEQVALSRTQVCGILREELFQGDAFHQSDTHIFIIMG 38
It.B-525                  MSSRPFLECTDENSEVCLALYRKSLKAIDLDHEGTHFDGQYPHIFVTLG 49
Dm.B-502                                                MVCEGTHFDGKIPHTFVIFG 20

G6PDH-A          ASGDLAKKKIYPTLWALFRDGLLPQKTKFVGYARTKMTLEELWAKIVPFLKVKDEEKSRF 101
G6PDH-B          ---------------ALFRDGLLPQKTKFVGYARTKMTLEELWAKIVPFLKVKDEEKSRF 51
G6PDH-C          ASGDLAKKKIYPTLWALFRDGLLPQKTKFVGYARTKMTLEELWAKIVPFLKVKDEEKSRF 120
G6PDH-D          ASGDLAKKKIYPTLWALFRDGLLPQKTKFVGYARTKMTLEELWAKIVPFLKVKDEEKSRF 95
Rn-515           ASGDLAKKKIYPTIWWLFRDGLLPEDTFIVGYARSRLTVDDIRKQSEPFFKVTPEERPKL 98
Hs.B-515         ASGDLAKKKIYPTIWWLFRDGLLPENTFIVGYARSRLTVADIRKQSEPFFKATPEEKLKL 98
It.B-525         ASGDLARKKIYPTLWWLYRDNLIPSNTHFIGYARSNTTIQDIKSKCEPYMKVKKGEEKLY 109
Dm.B-502         ASGDLAKKKIYPTLWWLYRDDLLPKPTKFCGYARSMLTVDSIKEQCLPYMKVQPHEQKKY 80
                               *:**.*:*. * : ****:  *: .:  :  *::*.   *.

G6PDH-A          AEFTRANSYLSGKYDESSDFVVLNKAMEKLEGNSGGNRMFYMALPPTVFQQVASNIKQHC 161
Rn-515           EEFFARNSYVAGQYDDPASYKHLNSHMNALHQGMQANRLFYLALPPTVYEAVTKNIQEIC 158
Hs.B-515         EDFFARNSYVAGQYDDAASYQRLNSHMNALHLGSQANRLFYLALPPTVYEAVTKNIHESC 158
It.B-525         EEFWKLNYYISGKYDVRSDFEKLNQEISQYEKGPIANRLFYLALPPSVFEDVTVLIKNTC 169
Dm.B-502         EEFWALNEYVSGRYDGRTGFELLNQQLEIMENKNKANRIFYLALPPSVFEEVTVNIKQIC 140
                  :*   * *::*:**   .:  **. :.  .    .**:**:****:*:: *:  *:: *
                                                  Substrate binding site
G6PDH-A          MSKQGWTRVVIEKPFGRDSQSSAELSNHLAALFEESQIYRIDHYLGKEMVQNLMAIRFAN 221
Rn-515           MSQTGWNRIIVEKPFGRDLQSSNQLSNHISSLFREDQIYRIDHYLGKEMVQNLMVLRFAN 218
Hs.B-515         MSQIGWNRIIVEKPFGRDLQSSDRLSNHISSLFREDQIYRIDHYLGKEMVQNLMVLRFAN 218
It.B-525         MAEKGWTRIIIEKPFGRDSASSQKLSDHLSSLFDEHQLYRIDHYLGKEMVQNLMTLRFGN 229
Dm.B-502         MSVCGWNRVIIEKPFGRDDASSQALSDHLAGLFQEDQLYRIDHYLGKEMVQNLMTIRFGN 200
                 ::  **.*:::*******  **  **:*:: ** * *:*********.******.:**.*
                                ↓Lys238
G6PDH-A          QIFGPTWNRNNIASIVISFKEPFGTQGRGGYFDNFGIIRDVMQNHLLQIMSLVAMERPVS 281
Rn-515           RIFGPIWNRDNIACVILTFKEPFGTEGRGGYFDEFGIIRDVMQNHLLQMLCLVAMEKPAS 278
Hs.B-515         RIFGPIWNRDNIACVILTFKEPFGTEGRGGYFDEFGIIRDVMQNHLLQMLCLVAMEKPAS 278
It.B-525         RIFGPVWNRDNIASIQITFKEPFGTQGRGGYFDEFGIIRDIMQNHLLQILTLAAMEKPAT 289
Dm.B-502         KILSSTWNRENIASVLITFKEPFGTQGRGGYFDEFGIIRDVMQNHLLQILSLVAMEKPVS 260
                 :*:.. ***:***.: ::*******:*******:******:*******:: *.***:*.:

G6PDH-A          TNAEDIRNEKVKVLKCVPPITMDNVVLGQYVGKPGGT-GEETQGYLDDPTVPKGSHTATY 340
Rn-515           TDSDDVRDEKVKVLKCISEVETDNVVLGQYVGNPSGE-GEATNGYLDDPTVPHGSTTATF 337
Hs.B-515         TNSDDVRDEKVKVLKCISEVQANNVVLGQYVGNPDGE-GEATKGYLDDPTVPRGSTTATF 337
It.B-525         IHPDDIRDEKVKVLKSVKTLTLNDVVLGQYVGNPEGE-GEAKIGYLDDPTVPAGSVTPTY 348
Dm.B-502         CHPDDIRDEKVKVLKSIEALTLDDMVLGQYLGNPQGTNDDARTGYVEDPTVSNDSNTPTY 320
                  ..:*:*:*******.:  :  :::*****:*:* *  .:   **::****. .* *.*:
                                ↓Arg357   ↓Lys366                  ↓Arg393
G6PDH-A          ATAVAYINNERWEGVPFILRCGKALNERKAEVRIQYKEVAGDLFAGESKRNELVLRVQPD 400
Rn-515           AAAVLYVENERWDGVPFILRCGKALNERKAEVRLQFRDVAGDIFHQQCKRNELVIRVQPN 397
Hs.B-515         AAVVLYVENERWDGVPFILRCGKALNERKAEVRLQFHDVAGDIFHQQCKRNELVIRVQPN 397
It.B-525         AAAVLRIDNERWDGVPFILKCGKALNERKAEVRIQFEDVPGDIFDGKAKRNELVIRVQPG 408
Dm.B-502         ALGVLKINNERWQGVPFILRCGKALNERKAEVRIQYQDVPGDIFEGNTKRNELVIRVQPG 380
                 *  *  ::****:******:*************:*:.:*.**:*  : :*****:****.
                                ↓Asp421    ↑Arg370
G6PDH-A          EAIYVKFMTKKPGMAFDIEETELDLTYGSRYKGMVMPDAYERLILDVFYGSQVHFVRSDE 460
Rn-515           EAVYTKMMTKKPGMFFNPEESELDLTYGNRYKNVKLPDAYERLILDVFCGSQMHFVRSDE 457
Hs.B-515         EAVYTKMMTKKPGMFFNPEESELDLTYGNRYKNVKLPDAYERLILDVFCGSQMHFVRSDE 457
It.B-525         EALYVKLMVKTPGMAFDMEETELDLTYGHRYENVKLPDAYERLILDVFCGSQMHFVRSDE 468
Dm.B-502         EALYFKMMTKSPGITFDIEETELDLTYEHRYKDSYLPDAYERLILDVFCGSQMHFVRSDE 440
                 **:* *:*.*.**: *: **:* ****  **:.  :************ ***:*******
                                ↓Arg487
G6PDH-A          LAEAWRIFTPLLHQIDKEKPKPVPYEYGSRGFKEADDLWKRAGFKFYGTYRWTKA----- 515
Rn-515           LREAWRIFTPLLHKIDREKPQPIPYVYGSRGPTEADELMKRVGFQYEGTYKWVN-----P 512
Hs.B-515         LREAWRIFTPLLHQIELEKPKPIPYIYGSRGPTEADELMKRVGFQYEGTYKWVN-----P 512
It.B-525         LSEAWRIFTPLLHEIEGNRVQPIPYTFGSRGPKEADELLKKHNFNYTGSYKWYK-----S 523
Dm.B-502         LREAWRIFTPILHQIEKEHIRPITYQYGSRGPKEADRKCEENNFKYSGSYKWHGGKAATS 500
                 * ********:**:*: :: :*:.* :**** .***   :. .*:: *:*:*
G6PDH-A          ---
Rn-515           HKL 515
Hs.B-515         HKL 515
It.B-525         NK- 525
Dm.B-502         NH- 502
```

**Fig. 2** Conceptual translation of *R. microplus* G6PDH-A, -B, -C, and -D and alignment with G6PDH from *R. norvegicus* (Rn, **NP_058702**), *H. sapiens* (isoform B, Hs.B, **NP_001035810**), *I. typographus* (isoform B, It.B, **AAX45784**), and *D. melanogaster* (isoform B, Dm.B, **NP_523411**). Residues of the NADP+ coenzyme binding site (GASGDLA) are identified in bold text and those of the substrate binding site/active site (RIDHYLGKE) are boxed. Residues involved in subunit dimerization are underlined, and residues that are involved in binding structural NADP+are identified by arrows with numbering relative to the Hs.B sequence. Identical amino acid residues within the alignments are denoted by (*), and similarity is denoted by (:)

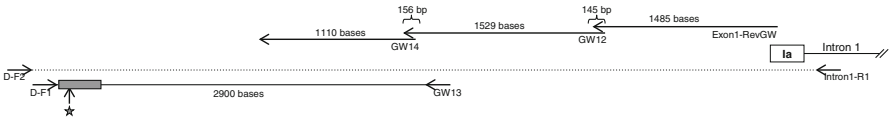



**Fig. 3** Primer walking and nested PCR strategy using *R. microplus* pooled genomic DNA template. This approach was utilized to isolate 5,300 bases upstream of the G6PDH-A translational start, as described in Results. Primer sequences are presented in Table 1. The *filled star* denotes the G6PDH-D translational start, and the *shaded box* represents the region that could not be amplified from gDNA template of individual *R. microplus* larvae M1 and M2 (see text)

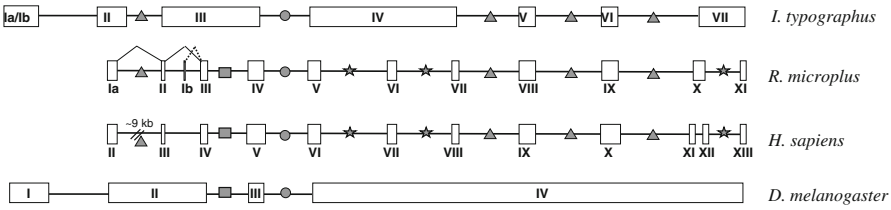



**Fig. 4** Comparison of gene structure for *R. microplus* (Rm) *G6PDH* with that of *I. typographus* (It), *H. sapiens* (Hs), and *D. melanogaster* (Dm). Exon sequences are depicted by a numbered box and are roughly drawn to scale. Spacing between exons, representing intron sequences, is not drawn to scale. Intron locations denoted by a *triangle* are shared by It/Rm/Hs, by a *square* are shared by Rm/Hs/Dm, by a *star* are shared by Rm/Hs, and by a *circle* are shared by all four

on M1 data (12,941 bases, Genbank Accession Number **EU595881**, Supp. Fig. 6) and only takes into account what is spliced to produce the G6PDH-A and -B transcripts since the start for G6PDH-D could not be verified within the M1 gDNA. The start of G6PDH-C within the gDNA sequence was not pursued in the course of this study.

The gene sequence spliced to produce G6PDH-A and -B is comprised of 12 exons (Fig. 4), the intron/exon boundaries of which have the characteristic GT and AG splice donor and acceptor recognition sequences (Breathnach and Chambon 1981; Supp. Fig. 6). The exons range in size from 17 to 218 bp, while the intron sequences vary from 172 to 929 bp. This genomic sequence data also includes 579 bases downstream of the translational stop codon, corroborating the 3′ UTR data obtained by 3′RACE.

The exon that determines the start of the G6PDH-A transcript was designated Exon Ia. The transcriptional start for G6PDH-B was localized within Intron 2 of this genomic sequence. This exon, designated Exon Ib, is 17 bases in length and is joined with Exon III and then IV thru XI to form the remainder of the G6PDH-B transcript (Fig. 4).

The *R. microplus G6PDH* gene structure is presented in Fig. 4 in comparison with available G6PDH gene structures from two insects (*D. melanogaster* and the bark beetle, *I. typographus*) and *H. sapiens*. The gene structure for *R. microplus* and human *G6PDH* have a striking similarity, extending to shared intron locations.

## Discussion

G6PDH is the rate-limiting enzyme in the production of NADPH, a role that is critical to lipogenesis and other biosynthetic reactions (Stryer 1995), and in the response to oxidative stress in prokaryotes (Storz et al. 1990) and eukaryotes (Ursini et al. 1997;

Filosa et al. 2003). Utilizing a tentative consensus sequence as a starting point, we have characterized three unique transcripts from *R. microplus* that encode G6PDH, designated G6PDH-A, -B, and -D, and identified a fourth transcript, G6PDH-C, from a salivary gland EST library (Olafson, unpublished).

The current, partially characterized gene structure for *R. microplus* G6PDH contains 12 exons (Fig. 4), and encodes the G6PDH-A and -B variants. In humans (Martini et al. 1986), cattle, fruit flies (Ganguly et al. 1985), and the bark beetle (Doleželová et al. 2006), G6PDH is encoded by a single gene. Sequence information obtained by amplifying pooled gDNA provided some evidence to support that G6PDH-D may be encoded by the same gene that encodes -A and -B. However, the inability to identify an ~840 base region that contains the G6PDH-D start from individual larval templates, coupled with the identification of three nucleotides at position 1,029 suggesting the duplicated nature of the G6PDH gene, indicate that the organization may be more complex. The prevalence of repetitive sequences within the *R. microplus* genome have been reported (~15.7% of the genome; Pagel Van Zee et al. 2007) and is evidenced by a 254 base indel that was identified in gDNA from both pooled and individual templates. Sunter et al. (2008) identified this indel sequence, named Ruka, as a novel small interspersed element that is present within the genomes of ixodid ticks, including *R. microplus*, in high copy number. Such repetitive elements may contribute to the complexity of gene organizations, and the presence of large indel heterozygotes between potentially duplicated copies of *G6PDH* may also play a significant role in compounding the gene characterization.

A comparison of the genomic organization for *G6PDH* between *R. microplus*, insects, and *H. sapiens* revealed that the tick structure is most similar to that of humans and, based on other published mammalian G6PDH gene structures, is also similar to that of *M. robustus* (wallaroo; Loebel and Johnston 1997), *Mus musculus* (mouse), and *T. rubripes* (pufferfish; Mason et al. 1995). A strong conservation of intron location between *G6PDH* of *R. microplus* and human was also observed, one minor exception being the intron separating exons XI and XII in the human sequence.

Single nucleotide polymorphisms (SNPs), a common source of genetic variation in eukaryotes (Brookes 1999), were observed within the various G6PDH-A, -B, -C, and -D transcripts isolated. Of the observed 35 SNPs within these transcripts, 97% were transition substitutions, which is in agreement with recent reports on the prevalence of transition polymorphisms in the Rocky Mountain wood tick (Alarcon-Chaidez et al. 2007) and *Anopheles funestus*, a vector of malaria (Wondji et al. 2007). Evaluation of SNPs within coding and non-coding regions may provide useful markers for tick population genetic studies, as has been accomplished in evaluating human populations (Seldin et al. 2006). With respect to *R. microplus*, eradicated from the US in 1960 (Graham and Hourrigan 1977), such population genetic studies would enhance epidemiological investigation of the geographic origins of ticks discovered on newly infested ranches within the permanent Quarantine or Buffer zone, a region established along the Texas-Mexico border to prevent re-establishment of the tick within the US (Bram et al. 2002). This is an important biological question since *Boophilus* sp. remain endemic in Mexico and the incidence of tick outbreaks within the Quarantine zone have steadily increased since 1999 (Bram et al. 2002).

Data from 3′RACE experiments of these four transcripts revealed two variant polyadenylated transcripts, denoted 3′-UTR-short and 3′-UTR-long. While a polyadenylation signal was identified in the 3′-UTR-short sequence (UAUAAA), this signal variant is only present in 4% of examined human 3′-UTR sequences (Chen et al. 2006) and 13% of *D. melanogaster* 3′-UTRs analyzed (Graber et al. 1999). The 3′-UTR-long sequence

contained both the non-canonical (UAUAAA) and canonical AAUAAA sequences, not unusual considering that multiple polyadenylation sites are represented in a significant fraction of human mRNAs (Gautheret et al. 1998). Graber et al. (1999) proposed that the presence of multiple polyadenylation signals may be used to regulate rates of polyadenylation and subsequently regulate gene expression. Whether such regulation is used to control expression of the described G6PDH transcripts is currently unknown.

Functionally important domains of G6PDH were observed in the conceptual translations of the isolated *R. microplus* transcripts, including the conserved substrate binding site/enzyme active site RIDHYLGKE (Jeffery et al. 1985) and the dinucleotide binding fingerprint GAxGDLx that has been associated with coenzyme NADP+binding (Levy et al. 1996). NADP+serves as a cofactor in the reaction catalyzed by G6PDH. In the *R. microplus* G6PDH-A, -C, and -D transcripts, both the enzyme active site and the catalytic NADP+binding site are completely conserved. G6PDH-B, however, encodes the enzyme active site but not the catalytic NADP+binding site, suggesting that the product may not be catalytically active. In mammalian G6PDH subunits there is also a separate NADP+molecule that plays a structural role (Au et al. 2000) and is integral for enzyme stability. Amino acid residues of importance for the structural NADP+binding site include four arginine, two lysine, and one aspartic acid residue, all of which are conserved within *R. microplus* G6PDH-A, -B, -C, and -D. In its active form, the G6PDH enzyme has been demonstrated to be present as a dimer or tetramer (dimer of dimers; Cohen and Rosemeyer 1969; Babalola et al. 1976), and the amino acid residues that play a role in the dimerization are located around positions 213–228 and 385–453 of the presented G6PDH-A conceptual translation and the respective residues for the G6PDH-B, -C, and -D translations.

The identification of more than one transcript encoding G6PDH in *R. microplus* was not unexpected, as two allelic products have been described for *D. melanogaster* (Giesel 1976; Fouts et al. 1988), *I. typographus* (Doleželová et al. 2006), and humans (Hirono and Beutler 1989). In *I. typographus*, the two alleles encode proteins that are 540 aa and 526 aa in length, and the transcriptional start encoding the smaller variant is located within an intron of the gene (Doleželová et al. 2006). This is similar to the location of the transcriptional start for *R. microplus* G6PDH -B within Intron 2 of the available tick genomic sequence. The presence of a third and fourth *R. microplus* transcript encoding G6PDH, though, is intriguing. In Arabidopsis, six G6PDH isoforms have been characterized that are encoded by five genes within the G6PDH gene family (Wakao and Benning 2005). The isoforms are functionally different with respect to tissue distribution and enzyme activity, and Wakao and Benning (2005) postulate that the different isoforms may play a combinatorial role in the response to variation in NADPH requirements, reflecting different metabolic requirements and photosynthetic activity within the organism. Expression analysis of the four *R. microplus* G6PDH transcripts indicated that both G6PDH-A and -C are detectable in all templates evaluated, suggesting these may be the major G6PDH isoforms in *R. microplus*. Of greater interest is the significantly elevated expression of G6PDH-A and -C that appears to be induced in fed females. This is in contrast to G6PDH-D which is undetectable in fed adult males and females, but has an elevated expression in unfed larvae. G6PDH-B was detectable, but its abundance was dramatically reduced in comparison to the other transcripts suggesting that it is a rare transcript and may lack catalytic activity due to the absence of the catalytic NADP+binding site. It has been observed that *R. microplus* adult females ingest approximately 100-fold their weight in blood during feeding (Sonenshine 1991), a process that is believed to be a source of general oxidative stress due to the release of heme subsequent to hemoglobin digestion. While the functional role of additional G6PDH transcripts in *R. microplus* is presently unclear,

it is plausible that engorgement induces the elevated expression of certain G6PDH isoforms in adult females to aid in the tolerance of feeding-induced oxidative stress by enabling an increase in NADPH production. Citelli et al. (2007) recently demonstrated that oxidative stress impairs the heme detoxification process in *R. microplus*. The oxidative stress was induced in their study by inhibiting the activity of catalase resulting in an increase of hydrogen peroxide, considered a reactive oxygen species. The formation of active catalase requires NADPH, supporting the possibility that the induced expression of G6PDH-A and -C in fed females but not fed males may correlate with an increased need for antioxidant defense, i.e. increased NADPH requirement, hence increased G6PDH that is critical to the function of NADPH-dependent enzymes involved in the general cellular response to stress. The difference of expression in adult males may be a function of bloodmeal volume and the time adult females spend on the host relative to adult males. In addition, the increased larval expression of G6PDH-D relative to its expression in other tissues/life stages analyzed and to larval expression of the other three transcripts may support a role for G6PDH-D in larval stage tolerance of oxidative stress.

The work described herein provokes further investigation into the role that G6PDH-A, -B, -C, and -D may play in enabling tolerance of oxidative stress throughout the life-cycle of a one-host tick, and raises the question of whether multiple G6PDH isoforms are present in other, multi-host tick species. A query of the EST database housed at NCBI identified a transcript from the three-host tick, *Amblyomma variegatum* (tropical bont tick; **BM292094**) that displays 83% similarity to the transcripts encoding G6PDH from *R. microplus*, providing an investigative starting point.

**Acknowledgments** The authors greatly appreciate the superb technical assistance of Ms. Greta Buckmeier, and thank Haiqi He, Andrew Li, and Pat Holman for their critique of an earlier draft of this manuscript. This article represents the results of research only. Mention of a proprietary product does not constitute an endorsement or recommendation by USDA for its use.

## References

Alarcon-Chaidez FJ, Sun J, Wikel SK (2007) Transcriptome analysis of the salivary glands of *Dermacentor andersoni* Stiles (Acari: Ixodidae). Insect Biochem Mol Biol 37:48–71

Au SWN, Gover S, Lam VMS, Adams MJ (2000) Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP+ molecule and provides insights into enzyme deficiency. Structure 8:293–303

Babalola AO, Beetlestone JG, Luzzatto L (1976) Genetic variants of human erythrocyte glucose-6-phosphate dehydrogenase. Kinetic and thermodynamic parameters of variants A, B, and A- in relation to quaternary structure. J Biol Chem 251:2993–3002

Bram RA, George JE, Reichard RE, Tabachnick WJ (2002) Threat of foreign arthropod-borne pathogens to livestock in the United States. J Med Entomol 39:405–416

Breathnach R, Chambon P (1981) Organization and expression of eukaryotic split genes coding for proteins. Annu Rev Biochem 50:349–383

Brookes AJ (1999) The essence of SNPs. Gene 234:177–186

Chen J-M, Férec C, Cooper DN (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. Hum Genet 120:1–21

Citelli M, Lara FA, da Silva Vaz I Jr, Oliveira PL (2007) Oxidative stress impairs heme detoxification in the midgut of the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. Mol Biochem Parasitol 151:81–88

Cohen P, Rosemeyer MA (1969) Subunit interactions of glucose-6-phosphate from human erythrocytes. Eur J Biochem 8:8–15

Davey RB, Garza J, Thompson GD Jr, Drummond RO (1980) Ovipositional biology of the cattle tick, *Boophilus annulatus* (Acari: Ixodidae), in the laboratory. J Med Entomol 17:287–289

deCastro JJ (1997) Sustainable tick and tick-borne disease control in livestock improvement in developing countries. Vet Parasitol 71:77–97

Doleželová E, Žurovec M, Böhmová M, Sehnal F (2006) Use of two transcriptional starts in the *G6PD* gene of the bark beetle, *Ips typographus*. Insect Mol Biol 15:25–32

Filosa S, Fico A, Paglialunga F, Balestrieri M, Crooke A, Verde P, Abrescia P, Bautista JM, Martini G (2003) Failure to increase glucose consumption through the pentose-phosphate pathway results in the death of glucose-6-phosphate dehydrogenase gene-deleted mouse embryonic stem cells subjected to oxidative stress. Biochem J 370:935–943

Fouts D, Ganguly R, Gutierrez AG, Lucchesi JC, Manning JE (1988) Nucleotide sequence of the *Drosophila* glucose-6-phosphate dehydrogenase gene and comparison with the homologous human gene. Gene 63:261–275

Gadallah AI, Khalil GM, Dees WH, Roshdy MA, Marzouk AS, Sonenshine DE, Main AJ (1990) Biochemical changes in *Hyalomma* (*Hyalomma*) *dromedarii* (Acari: Ixodidae) embryos and effect of 20-hydroxyecdysone applied to the mother. J Med Entomol 27:763–772

Ganguly R, Ganguly N, Manning JE (1985) Isolation and characterization of the glucose-6-phosphate dehydrogenase gene of *Drosophila melanogaster*. Gene 35:91–101

Gautheret D, Poirot O, Lopez F, Audic S, Claverie J-M (1998) Alternate polyadenylation in human mRNAs: a large scale analysis by EST clustering. Genome Res 8:524–530

George JE, Pound JM, Davey RB (2004) Chemical control of ticks on cattle and the resistance of these parasites to acaricides. Parasitology 129:S353–S366

Giesel JT (1976) Biology of a duplicate gene system with glucose 6-phosphate dehydrogenase activity in *Drosophila melanogaster*: genetic analysis and differences in fitness components and reaction to environmental parameters among Zw genotypes. Biochem Genet 14:823–833

Graber JH, Cantor CR, Mohr SC, Smith TF (1999) In silico detection of control signals: mRNA 3′-end-processing sequences in diverse species. Proc Natl Acad Sci USA 96:14055–14060

Graça-Souza A, Maya-Monteiro C, Paiva-Silva GO, Braz GRC, Paes MC, Sorgine MHF, Oliveira MF, Oliveira PL (2006) Adaptations against heme toxicity in blood-feeding arthropods. Insect Biochem Mol Biol 36:322–335

Graham OH, Hourrigan JL (1977) Eradication program for the arthropod parasites of livestock. J Med Entomol 6:629–658

Guerrero FD, Miller RJ, Rousseau ME, Sunkara S, Quackenbush J, Lee Y, Nene V (2005) BmiGI: a database of cDNAs expressed in *Boophilus microplus*, the tropical/southern cattle tick. Insect Biochem Mol Biol 6:585–595

Hill CA, Gutierrez JA (2003) A method for extraction and analysis of high quality genomic DNA from ixodid ticks. Med Vet Entomol 17:224–227

Hirono A, Beutler E (1989) Alternative splicing of human glucose-6-phosphate dehydrogenase messenger RNA in different tissues. J Clin Invest 83:343–346

Jeffery J, Hobbs L, Jornvall H (1985) Glucose-6-phosphate dehydrogenase from *Saccharomyces cerevisiae*: characterization of a reactive lysine residue labeled with acetylsalicylic acid. Biochemistry 24:666–671

Jongejan F, Uilenberg G (2004) The global importance of ticks. Parasitology 129:S3–S14

Kumar S, Thomas A, Pillai MK (1991) Involvement of mono-oxygenases as major mechanisms of delta-methrin resistance in larvae of three species of mosquitoes. Indian J Exp Biol 29:379–384

Lara FA, Lins U, Paiva-Silva G, Almeida IC, Braga CM, Miguens FC, Oliveira PL, Dansa-Petretski M (2003) A new intracellular pathway of haem detoxification in the midgut of the cattle tick *Boophilus microplus*: aggregation inside a specialized organelle, the hemosome. J Exp Biol 206:1707–1715

Lara FA, Lins U, Bechara GH, Oliveira PL (2005) Tracing heme in a living cell: hemoglobin degradation and heme traffic in digest cells of the cattle tick *Boophilus microplus*. J Exp Biol 208:3093–3101

Levy HR, Vought VE, Xiaohong Y, Adams MJ (1996) Identification of an arginine residue in the dual coenzyme-specific glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* that plays a key role in binding NADP+but not NAD+. Arch Biochem Biophys 326:145–151

Li AY, Davey RB, Miller RJ, George JE (2003) Resistance to coumaphos and diazinon in *Boophilus microplus* (Acari: Ixodidae) and evidence for the involvement of an oxidative detoxification mechanism. J Med Entomol 40:482–490

Loebel DAF, Johnston PG (1997) Analysis of the intron-exon structure of the *G6PD* gene of the wallaroo (*Macropus robustus*) by polymerase chain reaction. Mamm Genome 8:146–147

Luzzatto L (2006) Glucose 6-phosphate dehydrogenase deficiency: from genotype to phenotype. Haematologica 91:1303–1306

Martini G, Toniolo D, Vulliamy T, Luzzatto L, Dono R, Viglietto G, Paonessa G, D'Urso M, Persico MG (1986) Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. EMBO J 5:1849–1855

Mason PJ, Stevens DJ, Luzzatoo L, Brenner S, Aparicio S (1995) Genomic structure and sequence of the *Fugu rubripres* glucose 6-phosphate dehydrogenase gene (G6PD). Genomics 26:587–591

Moraes J, Galina A, Alvarenga PH, Rezende GL, Masuda A, daSilva Vaz I Jr, Logullo C (2007) Glucose metabolism during embryogenesis of the hard tick *Boophilus microplus*. Comp Biochem Physiol Part A 146:528–533

Pagel Van Zee J, Geraci NS, Guerrero FD, Wikel SK, Stuart JJ, Nene VM, Hill CA (2007) Tick genomics: the Ixodes genome project and beyond. Int J Parasitol 12:1297–1305

Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G, Belmont JW, Klareskog L, Gregersen PK (2006) European population substructure: clustering of Northern and Southern populations. PLoS Genet 2:1339–1351

Sonenshine DE (1991) The biology of ticks,vol 1. Oxford Press, New York

Storz G, Tartaglia LA, Farr SB, Ames BN (1990) Bacterial defenses against oxidative stress. Trends Genet 6:363–368

Stryer L (1995) Biochemistry. Freeman W.H, New York

Sunter JD, Patel SP, Skilton RA, Githaka N, Knowles DP, Scoles GA, Nene V, de Villiers E, Bishop RP (2008) A novel SINE family occurs frequently in both genomic DNA and transcribed sequences in ixodid ticks of the arthropod sub-phylum Chelicerata. Gene 415:13–22

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Res 22:4673–4680

Ursini MV, Parrella A, Rosa G, Salzano S, Martini G (1997) Enhanced expression of glucose-6-phosphate dehydrogenase in human cells sustaining oxidative stress. Biochem J 323:801–806

Wakao S, Benning C (2005) Genome-wide analysis of glucose-6-phosphate dehydrogenase in Arabidopsis. Plant J 41:243–256

Wondji CS, Hemigway J, Ranson H (2007) Identification and analysis of single nucleotide polymorphisms (SNPs) in the mosquito *Anopheles funestus*, malaria vector. BMC Gneomics 8:5–17

Yeh GC, Daschner PJ, Lopaczynska J, MacDonald CJ, Ciolino HP (2001) Modulation of glucose 6-phosphate dehydrogenase activity and expression association with aryl hydrocarbon resistance. J Biol Chem 276:34708–34713